

THE WORK THAT VANISHED – INAUGURAL LECTURE ON 28TH OCTOBER 2008

by Michael Seadle

1 Introduction

This lecture is based on a pun, or rather on a linguistic ambiguity. The word 'work' in English can mean both an intellectual product such as a book or journal, or even data from a scientific experiment. Work can also mean tasks or jobs such as those that librarians have today. This lecture involves vanishing acts for both.

Why have I chosen to speak in English? There are two reasons. The personal one is that English is my mother tongue and an important lecture ought not suffer from an imperfect accent and flawed grammar. The more scholarly reason is that, for better or for worse, English has become the dominant scholarly language in the world of library and information science (LIS). Works in English are read worldwide. Works in German are read in Germany, Austria, and Switzerland.

2 The Research Question

Humboldt Universität zu Berlin invited me here to transform the Institut für Bibliotheks- und Informationswissenschaft into an internationally competitive I-School like the School of Information at the University of Michigan or the Graduate School of Library and Information Science at the University of Illinois. Part of this effort involves persuading our students to follow international scholarly standards, such as offering a clear research question, an explicit and acceptable research method, and a literature discussion that puts their topic into the context of current scholarly discourse. My students will rightly and properly complain if I do not follow my own standards. My research question looks at the reason why I was brought here and what I hope to do. Specifically the question is what transformations have taken place in the LIS field that persuaded the university to hire me and what consequences do they have for practice, teaching, and research. The answer interrelates the two meanings of the word 'work', partly because fears about both jobs vanishing and digital documents vanishing grow from the set of transformations that have been taking place over the last thirty years.

3 The Research Method

My academic training was as an historian, but historians (at least in the US) are notoriously methodologically eclectic and I realized sometime not long after finishing my dissertation that I was in fact intellectually more akin to the cultural anthropologists who had so strongly influenced my own Doktorvater. Within academic anthropology I am essentially a follower of Clifford Geertz. Anthropology is an empirical discipline, but the data tend to be impressionistic rather than concrete. Geertz was well aware of this problem and addressed the problem of persuasion in a number of his works. For him the solution was essentially literary. He argues that it was neither "a factual look or an air of conceptual elegance" that mattered, but rather persuading readers of having truly "been there." (Geertz 1988) Cultural anthropologists are not novelists, of course, or merely academic journalists. How people use language matters greatly, particularly the linguistic distinction made by Ferdinand Saussure between the signifier and the signified. Such differences between word and meaning offer an empirical basis for recognizing social groupings, particularly within studies of contemporary cultures.

My research method has been, in effect, to live as a native among the tribe of librarians for the last thirty years without quite losing my perspective as an observer. This is hoary practice for ethnographers when it comes to exotic societies and it is increasingly common for ethnographers to observe their own cultures. Corporations in fact hire ethnographers today to help with a variety of external tasks (such as understanding their customers) and internal tasks (such as communicating between programmers and business operations). Libraries in the US hire anthropologists as well, as the well-received study by Nancy Foster and Susan Gibbons shows. (Foster & Gibbons 2007) The lab-books for

anthropologists are the notes they make based on their observations. Today these are not always written notes. They can be pictures, videos, voice recordings. Occasionally there are circumstances where notes cannot be made immediately, perhaps because the observer is too deeply involved in the event itself. Memory is one of the most treacherous (and most common) forms of note taking. Nonetheless it plays a key role in the selection of the information. Data plays a role where possible, and anthropologists of contemporary societies generally cite statistics and similar evidence. Nonetheless, as Geertz says, proof in anthropology is more than an assemblage of data. Footnotes help, verbatim texts help even more, details impress, numbers normally carry the day. But in Anthropology anyway they remain somehow ancillary: necessary of course, but insufficient, not quite to the point. The problem – rightness, warrant; objectivity, truth – lies elsewhere, rather less accessible to dexterities of method. (Geertz 1995) Persuasion, at least in this lecture, depends on painting a picture of the world that you the listeners believe, or at least do not entirely disbelieve. That must be in part a literary as well as scholarly achievement, and that is one of the ways in which my methodology fits with other disciplines in Philosophical Faculty I such as European ethnography, history and philosophy.

4 Scholarly context

For this lecture there are two scholarly contexts that matter. The most specific of these is the discourse over digital preservation that began with the articles by Anne R. Kenney and Lynne K. Personius about the "The Cornell / Xerox / Commission on Preservation and Access Joint Study in Digital Preservation". (Kenney & Personius 1992). While I was not part of the project team, I interacted with them on a daily basis. I will not go into detail here about all of the articles that followed on this topic, but I will mention a few key contributions, such as the research that Margaret Hedstrom and Clifford Lampe did on emulation (Hedstrom & Lampe N.d.). The published research revolving around LOCKSS and integrity-checking (Rosenthal, Robertson, Lipkis, Reich & Morabito 2005), and perhaps my own look at social models (Seadle 2006). In Germany the work of Ute Schwens and Hans Liegmann and of Reinhard Altenhöner is important for its emphasis on readability and usability — issues that clearly matter to the library community. (Schwens & Liegmann 2004, Altenhöner 2006)

The other scholarly context is broader. It has to do with fears librarians have about the future of their profession. We see artifacts of this context in research about the library as place, where the role of the building often seems bound with our identity as a profession. (Freeman, Bennett, Demas, Frischer, Peterson & Oliver 2005) We see this also in articles about joining library and computing centers under a single information-centered administration. (Bolin 2005) Repeated insistence that paper works will not go away have the tone of a defensive reaction against the incursions of the digital world. (Snowhill 2001) Concern about the new I-School model serving libraries may reflect the anxieties of those who feel their training is no longer as valued as it should be within the profession. (Dillon & Norris 2005)

5 Transformations

5.1 Overview

The establishment of the MARC (Machine Readable Cataloging) standard in the US, the availability of tapes with cataloging copy from the Library of Congress, and growth of OCLC cut the need for catalogers dramatically. The rise of purchasing programs and the growth in the importance of journal subscriptions cut out a substantial portion of the book-selection workload for subject specialists. The advent of small, specialized "boutique" digital libraries in this context mattered far less than the massive conversion of journal publication from paper to digital formats. The efforts of national level organizations like (in the US) the Center for Research Libraries to offer storage for a paper copy of journals shows one of the tendencies in research libraries to abandon paper and rely instead on digital formats.

5.2 First Transformation

The evidence we believe is generally the evidence that we see ourselves. I began working at the University of Chicago Library when I was 26 and was still writing my dissertation. At that time the cataloging department filled a space twice the size of the Senatssaal where we are now gathered, and most of the librarians sat almost cheek by jowl in small cubicles. The term "cataloging" in America includes both a formal description of a work and the assignment of subject headings, classifications and other intellectually significant parts of the metadata-creation process. In German terms, these librarians did work that many German Fachreferenten do still. I worked for the South Asia Library. Earlier it had seemed necessary to find specialists who could create cataloging copy in the roughly 25 literary languages with publications in India and Pakistan, but finding people for all those languages was impossible. The South Asia Library took the only reasonable solution of relying largely on computer tapes that carried machine readable (MARC) cataloging copy from the Library of Congress, plus the help of people like me who could learn enough of the languages to match copy when necessary. By the time I left that library after nearly 5 years for a career in computing, the crowding in the cataloging room had eased visibly. This was the first step toward transforming libraries into digital operations. A big piece of traditional work vanished. But another set of tasks appeared. The computing staff in the basement had visibly grown in numbers – almost as fast as the chairs in the cataloging area had emptied.

5.3 Second Transformation

Another piece of evidence came in the following decade. The most visible items of furniture on entering most research libraries were the row after solid row of cabinets for catalog cards. They were generally handsome objects of polished wood with steel or bronze-colored handles and label holders. This furniture held the most important tool for both librarians and researchers, because only via the catalog cards could anyone find particular items among the five million or so volumes in the stacks.

The timeframe in which Online Public Access Catalogs (OPACs) replaced this furniture with less lovely though highly utilitarian terminal clusters varied. At Northwestern University the process happened early. James Agard began programming on the NOTIS online catalog in the early 1970s. By the late 1980s essentially all serious research libraries had OPACS and automated systems. In Germany the elimination of card catalogs began later and proceeded slowly. A few libraries still have card catalogs for their older materials. The conversion from paper to digital record keeping is, however, now largely an accomplished fact and with it a major transformation has occurred. Today the only way for librarians and researchers to find a particular work is to interact with a computer system. The old task of filing cards into the catalog vanished totally. Whole departments of filing specialists closed or redistributed their staff to new tasks, in so far as their training allowed.

5.4 Third Transformation

The third transformation is contemporary but no less obvious. I noticed it first when I began to hear that new faculty at Michigan State University's Business School no longer just asked whether the library subscribed to particular journals, but asked whether it had those journals in electronic form. My own publisher, Emerald, had long since digitized back issues. The paper copies continue to be printed for journals that traditionally had paper versions, even though our usage statistics today rely entirely on the digital copies. Increasingly libraries today choose digital-only options for subscriptions, or make a deal in which only one library in a consortium keeps a paper copy. The effect is that current journal acquisitions have become increasingly digital. Current periodical reading rooms in US research libraries have shrunk noticeably because they have fewer paper copies to display. With this change the old job of checking-in physical copies of journals has diminished, as has the work of gathering the fascicles for binding at the end of the year. Most of these were low level jobs in areas that the public rarely saw. The elimination of paper journals offered an opportunity to address the space problem from ongoing acquisition of paper-based monographs. In order to make room for these

books, libraries began sending paper versions of digitized journals (particularly those in the JSTOR collection) to remote storage, or discarding them entirely. New reading tools like the iRex and Sony Reader may make it more likely that monographs will follow scholarly journals in the shift to digital formats. Whether paper survives in some form or not is irrelevant. The changes in journal subscriptions have already transformed the library world.

6 Consequences

6.1 Overview

'Work' in the form of mechanical jobs has vanished and new work in the form of technology tasks has appeared. 'Works' in paper are vanishing and works in digital formats have taken on a major role in modern scholarship. The next section looks at the consequences for library practice, research, and teaching.

6.2 Practice

Specific consequences of recent transformations can be found in the staff directory and the staffing budget for US research libraries. The number of names in the staff directory at US research libraries has tended to decline as poorly paid low level positions vanish, but staff budgets continue to grow because libraries are replacing people who did routine jobs with better-paid, better-educated professionals. It is clear from job advertisements that these new librarians need to have skill-sets that include a strong technology background. Employers typically also request an ability to work in teams and a willingness to deliver a useful product to the end-user. These are skills that businesses need too and unsurprisingly businesses have been hiring many of the best students from US I-Schools.

Librarians without such skills can compete at best for low-level positions. When some jobs vanish, others appear, but when a document vanishes, it may be gone forever. Libraries were among the first institutions to raise the alarm about the vulnerability of digital works to loss over time, especially those institutions like Cornell that already had strong technology skills. This concern has spawned a number of major digital archiving projects with most major US research libraries financially supporting one or more, particularly LOCKSS and Portico. Relatively few German libraries have incorporated serious long term digital archiving into their routine operations. It would be unfair to say that they do not worry about digital works vanishing, but many prefer to wait for a right solution to appear. The right solution may, however, appear too late, and it is hard to know what a good solution is without testing multiple possibilities. KOPAL has evoked substantial interest and a German LOCKSS community is being built. This is progress, but much fundamental research in this area remains to do. The library community has, for example, little or no experimental data to project how reading habits and needs might evolve over time. The assumption tends to be that people in 100 years will read as they do today, even though we know that readings habits and methods 100 years ago were different. Research in this area is urgently needed.

6.3 Research

I have heard librarians in Germany say that librarians ought not do research and should merely serve the needs of other researchers. At the APE conference in January 2008 in Berlin, Peter Murray-Rust from Cambridge asked whether the work that librarians did in providing journal subscriptions was any different than the work a purchasing once did in ordering chemicals. No librarian could offer an answer that he found convincing. Librarians that only buy and deliver publications are in his terms clerks, not professionals. One area where librarians ought to engage in serious research is long term digital archiving. Librarians have long had responsibility for ensuring that information persists. Much of the current discussion on long term digital archiving has focused on technology problems such as how to ensure the integrity of the bitstream and how to make reasonable provision for the usability of a digital object in 200 years. These are important, but the issues are not merely technological. For example archiving systems today tend to define the integrity of a digital object by calculating its

check-sum. The check-sum is a reasonable integrity check at one level, since any change in the digital document whatsoever will (under ordinary circumstances and with a very high degree of likelihood) change the check-sum. But this is a purely mechanical integrity measure. Few people would say that the integrity of a work were lost because someone highlighted a word on a page, even though the act of highlighting changes the check-sum. A useful research question might be: how to define integrity in computing terms so that the meaning maps well to socially accepted norms. Today people worry that digital works will vanish if their formats cannot be migrated to up-to-date versions. The library community is attempting to persuade the non-library world to adopt preservation-friendly formats. This is a logical solution, but the library community has had little success in persuading corporations to adopt its standards, except for products sold directly to libraries. We also know that most files can be reverse-engineered with sufficient time and tools. The barriers are time and cost, not capability. Research into the costs and tools could provide alternatives if preservation-friendly formats fail to become standard.

A related research project is to ask to what degree preserving the look and feel of current documents matters for the long term. In a small experiment with students last semester my assumption that the visual context would matter to them was overturned. Project Gutenberg has preserved older works in plain ASCII text. Plain ASCII also loads relatively easily into ebook readers like the iRex or Sony. Newly published editions of nineteenth century books in fact almost always reformat the text to modern fonts and layouts. It may be that our assumptions about context preservation need testing. Part of the testing needs also to consider how humans interact with works. In one of my classes I talk about the evolution of the "book" as an information system. We know that people have interacted differently with books over time as their structure changed and as the number and variety of books grew. It seems likely that these relationships will continue to evolve and that the interactions between humans and computing systems may also become a key area of research for understanding the long-term usability of archived information.

6.4 Teaching

Leading North American library programs have replaced courses in well-established subjects like cataloging with a curriculum that includes human-computer-interaction and information economics. At first libraries doubted whether they wanted people with that kind of training. Ten years later employment for students from these programs is near 100 per cent and jobs for students from more traditional schools that emphasize so-called practical skills are harder to find. Success in the modern library world depends not just on technological facility, though technology is important. Library schools recognize that students need the ability to solve problems and adapt to new circumstances. Skills of that sort come from engaging with real questions. This is one reason why I-Schools try to involve students in cutting edge research where the students learn that professors do not have the answers: instead they have research methods that can be applied to get answers. At Humboldt we have a proud tradition of the unity of teaching and research. This tradition fits well with plans for the Institut für Bibliotheks- und Informationswissenschaft, but it also means shifting our teaching from more traditional content. Students make particularly good subjects to involve in research on long term digital archiving because they are in some sense the first generation that really will have to rely on effective digital archiving when they reach my age. Interactions with them have already changed my ideas about issues like readability and data integrity.

7 Conclusions

When I accepted Humboldt's invitation to take this professorship, I mentally also accepted the challenge to do what I could to prepare our students for a future that some of them are not quite sure that they want. One of our retired faculty has said that when students used to inquire about becoming a librarian because they loved books and reading, she told them to study something else. We may nonetheless be a profession of 7 book lovers, but if the transformations of the last 30 or 40 years mean anything, our job is something else. A bitstream is probably inherently less lovable than a handsomely-bound book. Nonetheless a bitstream appears to be the future of information and we

need to prepare our students for dealing with it. The challenge that libraries face in making sure that these works do not vanish demands hard thought and rigorous research. The work that vanished could still vanish. We could also vanish as a profession if we cling to the past, and we could lose a significant part of the intellectual output of the present day if we lock ourselves into solutions without empirical research about the probable effectiveness of those solutions in a century or two. Nonetheless libraries are fairly robust institutions. They have survived many transformations over the centuries and odds are they can survive again with their contents intact if our students make the right choices. The choice, dear students, is up to you.

References

- Altenhöner, Reinhard. 2006. "Data for the future." *Library Hi Tech* 24(4):574–582.
- Bolin, M.K. 2005. "The Library and the Computer Center: Organizational Patterns at Land Grant Universities." *The Journal of Academic Librarianship* 31(1):3–11.
- Dillon, A. & A. Norris. 2005. "Crying Wolf: An Examination and Reconsideration of the Perception of Crisis in LIS Education." *JOURNAL OF EDUCATION FOR LIBRARY AND INFORMATION SCIENCE* 46(4):280. URL: <http://www.ischool.utexas.edu/~adillon/Journals/JELIS.pdf>
- Foster, Nancy Fried & Susan Gibbons. 2007. *Studying Students: The Undergraduate Research Project at the University of Rochester*. Chicago, IL.: American Library Association.
- Freeman, G.T., S. Bennett, S. Demas, B. Frischer, C.A. Peterson & K.B. Oliver. 2005. "Library as Place: Rethinking Roles, Rethinking Space. CLIR Publication No. 129." Council on Library and Information Resources p. 89. URL: <http://www.clir.org/pubs/abstract/pub129abst.html>
- Geertz, Clifford. 1988. *Works and Lives: the Anthropologists as Author*. Stanford, Ca.: Stanford University Press.
- Geertz, Clifford. 1995. *After the fact: two countries, four decades, one anthropologist*. Cambridge, MA, Harvard University Press.
- Hedstrom, Margaret & Clifford Lampe. N.d. "Emulation vs. Migration: Do Users Care?" RLG Diginews. Forthcoming.
- Kenney, Anne R. & Lynne K. Personius. 1992. *The Cornell / Xerox / Commission on Preservation and Access Joint Study in Digital Preservation*. Technical report Commission on Preservation and Access.
- Rosenthal, D.S.H., T.S. Robertson, T. Lipkis, V. Reich & S. Morabito. 2005. "Requirements for Digital Preservation Systems: A Bottom-Up Approach." Arxiv preprint cs.DL/0509018. URL: <http://arxiv.org/abs/cs/0509018v1>
- Schwens, Ute & Hans Liegmann. 2004. *Langzeitarchivierung digitaler Ressourcen*. In: *Grundlagen der Grundlagen der praktischen Information und Dokumentation*. Saur. pp. 567–570.
- Seadle, M. 2006. "A Social Model for Archiving Digital Serials: LOCKSS." *Serials Review* 32(2):73–77.
- Snowhill, L. 2001. "E-books and their future in academic libraries." *D-Lib Magazine* 7(7/8):1082–9873. URL: <http://www.dlib.org/dlib/july01/snowhill/07snowhill.html>